DOCUMENT RESUME

ED 282 466 HE 020 198

TITLE Research and Development Update.
INSTITUTION College Entrance Examination Board, New York, N.Y.
PUB DATE Jan 87
NOTE 7p.; Papers presented at the National Forum of the College Board (October 1986).
PUB TYPE Collected Works - Serials (022) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Standards; *Accountability; *Educational Assessment; *Educational Quality; Higher Education; Instructional Effectiveness; *Outcomes of Education; School Effectiveness; Student Development; *Student Evaluation; Teacher Role
IDENTIFIERS *Value Added

ABSTRACT

College outcomes evaluation is discussed in two papers. In "The Political Economy of Assessment," Joseph P. O'Neill considers standards of accountability, American reliance on faculty to assess, and the danger of validating quality by external examination. By using the credit-hour system, a quantity indicator (time), is used as a standard. There is no mechanism external to the classroom to verify the integrity of the baccalaureate degree; the faculty member teaches, examines, and certifies. It is noted that large classes and extensive use of graduate students and part-time faculty have necessitated outside validation of quality. Five axioms important to the successful assessment of student progress in college and the complexities of measuring "value added" are considered by William W. Turnbull in "Can 'Value Added' Add Value to Education?" It is suggested that assessment cover many dimensions of student growth and learning and that a variety of assessment techniques be used, along with a baseline measure of the skills that the student has on entry. A way to assess the accomplishments of graduating students and to express the results in terms that have meaning from year to year or across departments or institutions is proposed. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RESEARCH AND DEVELOPMENT UPDATE

January 1987. Research and Development, The College Board, 45 Columbus Avenue, New York, New York 10023-6917

ED282466

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

College Board

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 020 142

BEST COPY AVAILABLE

Research and Development Update.

January 1987. Research and Development, The College Board, 45 Columbus Avenue, New York, New York 10023-6917

This issue includes two papers about higher education outcomes evaluation, which were delivered at the College Board's National Forum in October 1986. Taken together, the two provide the reader with policy and practical considerations in assessing higher education outcomes and a description of the changes within higher education that make external modes of assessment a useful, or possibly necessary, validator of quality. Although the authors support the assessment of educational outcomes of college, both caution that the concept of "value added" tends to oversimplify the complexity of the tasks involved. If assessment is to play a dynamic role in improving institutions it must be grounded in the history and operational modes of higher education institutions.

Research and Development Update, an occasional publication devoted to topics of current educational concern, provides researchers an opportunity to place data and their interpretation before an interested and concerned audience. Therefore, points of view or opinions expressed do not necessarily represent official College Board policy.

Robert G. Cameron, Executive Director
Research and Development

THE POLITICAL ECONOMY OF ASSESSMENT

JOSEPH P. O'NEILL

Reform a university? You may as well reform a cheese—there is a certain flavor about a university as there is about a cheese, springing from its antiquity, which may very easily be lost by mishandling.

—Lord Hugh Cecil

Though my topic is the political economy of assessment, I am not an economist. I produce manuals for administrators on college financing and facilities that most economists would greet with a polite yawn. I, for my part, view economics in much the same way that the ship's cook on the Beagle must have eyed Darwin's specimens—not as stepping stones to a grand theory, but for their practical use in the stew pot.

Yet, as someone who has peeked and poked into the economic underpinnings of a great many liberal arts colleges, I have begun to notice patterns of behavior that have to do with economy beyond its narrow financial meaning.

I am using the term "political economy" today as Adam Smith would have used it—that is, as the branch of moral philosophy that examines the value system that governs the ways in which a society distributes goods and services among its citizens. What, you may ask, has educational assessment to do with political economy? A great deal, as a matter of fact. Examinations, the most common form of assessment, ration access to such scarce goods as admission to our most selective colleges and graduate schools, to civil service employment, and to the practice of law and medicine.

My focus, however, is more on institutions than on the individual. A few weeks ago, for example, the governors of our 50 states issued a report, *Time for Results*,[1] that would "hold institutions accountable for their students' performance." At least one state, Colorado, has decreed that if its colleges do not implement an assessment program within five years, they will lose 20 percent of their state funding. Now, that's getting serious. And this concern about measuring outcomes is not confined to politicians. Even Derek Bok, the president of Harvard, devoted his 1984-85 report to the Board of Overseers to the question of assessment.

Why this sudden, seemingly universal interest in assessing the outcomes of a college education? Why have assessment and different forms of examinations become the center of attention? In the Sputnik era, when we had similar concerns about education, we focused on reform of the curriculum, not examinations or assessment. If we examine past academic reform movements that stressed the evaluation of outcomes—Oxford and Cambridge in the mid-nineteenth century or the Flexner Report and its effects on medical education in the United States in the early years of this century—we find that examinations are typically a vehicle to reform not so much the student body as the teaching body. If the nation's governors are serious in their demand that institutions be held accountable for the performance of their students, then college administrations have little choice but to hold the faculty accountable for that same performance.

Joseph P. O'Neill is Scholar in Residence at the Carnegie Foundation for the Advancement of Teaching.

1. *Time for Results: The Governors' 1991 Report on Education*. Washington, D.C.: National Governors' Association Center for Policy, Research, and Analysis, 1986.

Standards of Accountability

Accountability, not assessment, is the real issue. If we compare our standards of accountability to those of other professions, we find that ours are seriously fragmented. In medicine, law, accounting, or engineering there is always an attending physician, a lead attorney, or a partner-in-charge who is responsible to the client for whatever work is done. Maim a patient, lose a case, or botch an audit, and there is no question who is responsible. Undergraduate teaching, in contrast, has become a profession in which all are responsible, but none can be held accountable for overall student performance. As Bok put it, "Even aims acknowledged by all, such as the ability to communicate precisely or to reason carefully, may remain the responsibility of no one."

What are the mechanisms that dilute accountability and hide from us the true quality of undergraduate education? The first and most universal of these is the credit-hour system. In assigning credits to units of instruction we use a standard of account that measures only a quantity, *time*, but not quality. A three-credit course taught by a Nobel Laureate or by a part-time pickup is counted equally toward a degree, not because of the quality of instruction but because 37½ hours of class time went into it. No matter how well or poorly a subject is taught, the credit hours taught by the faculty and consumed by students are considered to be homogeneous. Since credit is denominated by a quantity of time with no reference to quality, colleges in hard times are tempted to choose the less expensive instructor or to pack the hall—or, in the worst of both worlds, to do both at the same time. Like governments that devalue their currency, colleges can debase the quality of instruction while maintaining the nominal credit value for courses. Large-scale use of graduate students and part-time faculty in undergraduate education has become the collegiate equivalent of clipping coin.

But the credit-hour system is merely a symptom, not a cause. If we examine our current structure of teaching and evaluation, we may find that there is a conflict of interest in the way that American colleges and universities certify instruction—that is, grant credit toward a degree. In the American system, there is no arms-length relationship, at the undergraduate level at least, between the teaching function and the certifying function. Faculty members not only teach but in effect guarantee, first, that their teaching meets established standards in both content and quality and, second, that students have learned what the faculty have taught. There is no mechanism external to the classroom to verify the integrity of the baccalaureate degree. We are so accustomed to the conjunction of the teaching and certifying (i.e., credit-granting) functions in the individual faculty member that even the mention of separation might seem exotic. Yet, this American practice is by no means universal.

If we examine the customs of Oxford and Cambridge, we find that the colleges teach and the university examines and grants degrees. An Oxford college establishes its reputation within the university by the number of "firsts" that its graduates receive in university-wide examinations. One American Rhodes Scholar on his return from Oxford remarked: "I remember that Magdalene College felt disgraced because it did not obtain a single first class one year." Concern about the quality of teaching and learning at Oxbridge colleges is more acute than with us, because the tutor as well as the student will be evaluated by a jury of peers.

American Reliance on Faculty to Assess

I do not say that the English or European systems of external examinations are better than our course-by-course evaluations. Indeed, as I shall explain later, our reliance on the integrity of the faculty rather than external examinations has its unique strengths. Instead, I am exploring the unanticipated consequences of the path we have chosen.

If we compare the patterns of teaching and the patterns of faculty authority before and after the Civil War, we find a gradual change from the recitation to the lecture. At the same time the professor's classroom authority was totally transformed. Our current classroom tradition of pop quizzes and frequent intermediate tests is a direct descendant of the recitation method in the nineteenth-century liberal arts college. Abandoned in the latter part of the nineteenth century was the standard and almost universal practice, even at Harvard, of outside examiners' questioning students in public oral examinations and thereby validating the quality of the college's undergraduate education. Charles Eliot, in his inaugural address as president of Harvard in 1869, called for an external examining body that would be distinct from the teaching body in the granting of degrees. "When the teacher examines his class," said Eliot, "there is no effective examination of the teacher." Eliot was to remark in his later years that this was the one recommendation in his inaugural address that had not been implemented.

By the beginning of the twentieth century a professor's classroom authority was unquestioned. In its 1915 General Declaration of Principles, the Association of American University Professors (AAUP) compared the authority of a faculty member with that of a federal judge. "University teachers," said the AAUP statement, "should be with respect to the conclusions reached and expressed by them no more subject to the control of a board of trustees than judges are to the control of the President with respect to their decisions."

The self-contained course had by that time become the centerpiece of American higher education. In it reigned the self-verifying faculty member who teaches, examines, and certifies—that is, grants credit toward a degree without external validation. The original model for this all-encompassing faculty role was the *Lehrfreiheit* of the German university professor. A Mommsen or a von Harnack was a scholar without peer teaching what and how he wanted, because no one else was competent to question his authority.

But this professorial authority and independence had,

Copyright © 1987 by College Entrance Examination Board. All rights reserved. College Board, Scholastic Aptitude Test, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board.

to many minds, its dark side. When asked to describe the difference between a German university professor and his American counterpart, James Morgan Hart, who had studied at Göttigen and Berlin in the 1860s, said: "[In Germany] the professor is not a teacher, in the English sense of the term; he is a specialist. He is not responsible for the success of his hearers. He is responsible only for the quality of his instruction." In such an ethos, any demand that the faculty be held responsible for student performance would have been incomprehensible. And to the extent that the German model of the university professor has prevailed, especially in our research universities and large comprehensive institutions, some measure of that incomprehension continues. In a recent survey by the Carnegie Foundation, 45 percent of students enrolled at large comprehensive universities said they had no personal contact with faculty members.

The current advocates of assessment would change all that. If we look closely at what the process would require, we find that it would demand not only a transformation of the student–faculty relationship but new forms of record keeping, a professional staff to convert that new data into usable information, and, above all, a willingness on the part of faculty to make use of the information gained from assessment to improve teaching and learning. Colleges that are serious about assessment have appointed specialists in evaluation, professionals whose task it is to interpret to the faculty the *practical* meaning and significance of what the assessment process has discovered. Their function is analogous to that of the specialist who interprets a CAT scan or a blood test to an attending physician. We would find it scandalous if a doctor were to ignore such tests or not incorporate them into the treatment of his patients. Yet in much of American higher education we rarely find organized, college-wide systems of evaluation that would feed back to the faculty the effect that their teaching has on students—not to speak of using such feedback to improve the quality of teaching and learning.

The Danger of Validating Quality by External Examination

There are clear caveats, however, to this new emphasis on assessment. Is it feasible, or even desirable, for faculty members to be responsible for their students in the *same* way that a physician or attorney is to a client? No, not, at least, if the passivity of the patient or the client's dependence on the attorney's judgment is emphasized. Students are apprentices who learn by doing. Their ultimate goal is to have the same kind of independent judgment and skill that their professors exhibit. The real danger of the current emphasis on assessment is that we may, without consciously realizing it, be replacing the "apprentice" model of education with a "medical" one. The paradox of education is that no one can be forced to become an independent learner.

Let me pull together the threads of my argument. In adopting the German model of professorial independence, we left behind the use of external examinations as checks and balances to that independence. By and large we Americans were the better for it, more flexible and inventive in our teaching and learning. But we could afford to abandon external validation of our degrees only as long as we protected the integrity of the professoriat. As long as we had reasonable student–faculty ratios and applied rigorous standards as to who would be allowed to teach in the classrooms, a formal system to check and balance professorial indepenence was largely unnecessary. However, once fiscal pressures created the auditorium-sized lecture room and forced the substitution of part-timers for regular faculty—a substitution veiled and facilitated by the credit-hour system—then comprehensive examinations or other modes of assessment that reach beyond the individual course have become once again a useful, perhaps necessary, validator of quality.

CAN "VALUE ADDED" ADD VALUE TO EDUCATION?

WILLIAM W. TURNBULL

The term "value added" is among our most abused buzzwords in education. It has been in the economists' vocabulary for a very long time. As you know, they use it to represent the extent to which a particular stage in the development or manufacture of a product enhances its worth. Examples would be the increase in the value of a raw diamond when it is cut and polished or the value added to the clay when it is shaped into a pot. At any rate, it was easy for the term to slip into educational parlance to represent the beneficial change in students brought about by the efforts of those responsible for their schooling, as the faculty polishes up the diamonds in the rough who enter their institution.

However, when the economic model is applied to education, use of the short-hand term "value added" is unfortunate. Unlike the manufacturer, diamond cutter, or potter, the educator is not dealing with inert material to be reshaped. Education is the process of fostering students' growth and beneficial change. The faculty role is more like that of the gardener who is trying to promote healthy growth in the garden plot. But, regardless of the precise term "value added" or the imperfect analogy in its application to education, the root idea of assessing how much students *learn* or *improve* or *grow* in school or in college, as well as how they stand at graduation, is not only a good and important idea but obviously one that lies near the heart of the education enterprise.

Progress versus End Product

Judging the quality of an institution's curriculum or teaching only by the final level of attainment of its graduating class is as illogical as it is common. I had a friend who used to say that he was going to start a school for producing tall people.

William W. Turnbull is Distinguished Scholar in Residence at Educational Testing Service, Princeton, New Jersey.

with a money-back guarantee that they would reach a height of 6 feet by the time of graduation at age 17. His only admission criterion at age 13 was that each student be at least 5 feet 10 inches tall.

For years, people have complained that the more selective institutions are like the school for producing tall people. They require students to be highly developed when they arrive and then have nothing to do but wait and receive the nation's applause when those same students emerge as exceptionally well-educated people. President Eliot of Harvard was once asked how the institution had managed to accumulate such impressive resources over the years. He is alleged to have responded, "It's simple: the students bring so much with them when they enter and take so little with them when they leave." At any rate, some people contend that there are less selective colleges to which most students come with much less impressive prior skills but in which they learn as much and progress as fast as the people in the more selective institutions. The colleges that produce those excellent gains, and their students, the argument goes, deserve the accolades of the public at least as much as those institutions that are blessed with the shiniest freshmen, who in due course emerge as the most lustrous graduates.

The counterargument is that it's the end product that counts. Some parents are infuriated with the value-added assessments of their children ever since kindergarten days: report cards saying, "This child shows improvement," but saying nothing about whether Johnny or Jane has suddenly begun quoting Shakespeare or merely stopped biting the teacher's ankles. They say you don't pick a lawyer or a brain surgeon for how much they improved while in school, but for how good they are at the end.

The fact is, of course, that both progress and end product are important. Which is more important depends on your purpose. Some measures of student progress can tell you how well both the institution and the student did with what they had to start with. Assessing the amount of improvement or progress made by students is most likely to be pertinent when one is trying to judge the success of one program of instruction compared with another. How well is it bringing a group of students along, compared with the program of another college or perhaps compared with the curriculum the college was following last year? Some measure of the standard reached by the end of the line is essential before a person can be certified as ready to enter practice or training that presumes a certain level of prerequisite knowledge or skill. Neither a measure of progress nor a measure of end status by itself is likely to tell you all you want to know.

Complexities of Measuring Value Added

One of the problems is that posing questions about progress or gains and especially about "value added" seems to drive people immediately to oversimplified approaches. For example, in the school for producing tall people, why not just measure every student at entrance and again at graduation, subtract the entry level from the exit level figure, and grade the success of the school by the average increase?

First, you have to know how many students have left between entrance and graduation. The school or college that keeps only the most successful students has a built-in guarantee that the graduating class will look good. (Even an open-door school for producing tall people could seem to show great results by flunking all the short students.) The fact that the average score of the people who are left to graduate is substantially higher than the average score of the entering freshmen may say more about its retention policy than about its instructional prowess. It's even worse if you try to infer change from a simple comparison of how much today's seniors in your college know, or their attitudes, compared with today's freshmen. I'm indebted to Ernie Pascarella for the observation that if you were to judge from *cross-sectional* age samples of the population of Miami you might conclude that most people there are born Cuban and grow old Jewish.

Another oversimplification relates to the kind of measure used. Where you are dealing with a criterion as one-dimensional as height and a test as simple and self-evident as a measuring tape, a before-and-after measure makes sense, as does subtracting the measure at the beginning from the one at the end. But suppose your aim goes beyond increasing height to improving health. Suddenly you have a whole array of measures that various people think are critical. You also have all the arguments about the relative importance of indicators such as weight, blood pressure, pulse rate, or cholesterol level and so on through the whole rich panoply of criteria of both fitness and wellness. When you broaden the criterion, the idea of subtracting a set of beginning measures from a set of end measures begins to sound inadequate for many reasons. It's the patterns and interrelations among the indicators that count. Moreover, by graduation time from a training program for athletes, you may be using tests of performance or endurance that would have been unthinkable, if not fatal, if given at the entry level. Bob Pace points out that a skiing instructor wouldn't send all his students down a steep hill on Day One so that he could demonstrate value added at the end of the course. The same is true in an academic institution. By the end of the program students should be dealing with levels of complexity that could not reasonably have been introduced into entry-level exercises. Accordingly, there is unlikely to be an index of academic attainment that can yield to simple subtraction to show the growth of college-level competencies.

For these and other reasons, I think that the term "value added" conjures up a simplistic and unidimensional cookie-cutter view of educational progress that is a potential trap for the unwary. And yet I do not believe that measuring status by itself is sufficient, and I am an enthusiastic advocate for assessing student gains in a broader sense, freed from the expectation that it can be done with the simplicity and precision often implied by "value added."

Assessing Student Progress in College

Five axioms are, I think, pertinent to the successful assessment of student *progress* in college.

- First, the assessment must be multidimensional. Every college that I am familiar with aspires to help its students gain in clear thinking, intellectual curiosity, broadened understanding, the ability to express oneself, career or vocational readiness, and knowledge both general and specific to a field of study, not to mention personal integrity and citizenship.
- Second, there just isn't any way to measure those attributes, simply or with precision, for a whole class—much less for an individual student.

A variety of assessment techniques—including passive indicators such as using the library or engaging in vandalism, or dropping out or succeeding in landing and holding a job—will need to be made part of an organized assessment program along with work samples such as the senior thesis or the departmental examination. As you might expect, I firmly believe that external tests can be helpful in assessment, but the idea that a test is going to give you more than a fraction of what you are interested in learning about progress toward the broad goals of higher education is, at this date, *totally illusory*.

- Third, when progress or growth is to be assessed, some way must be found to establish a baseline measure of the skills with which the students enter the institution. Even though it's seldom possible to have a plausible set of tests for incoming freshmen that is a reasonable replica of the multitude of skills you want to be able to observe at graduation, there is a reasonable alternative available: making use of the SAT scores of the incoming class to serve as a rough measure of developed ability at entrance.

For knowledge and skill objectives, in particular, it is possible to set up a set of expectancy tables showing the mean level of senior-year accomplishment, as measured by grades or test scores, to be expected of students who enter with a given level of proficiency as measured by the SAT. It is obviously much more difficult to achieve a high level of end-of-college performance with a group of students who entered college with an average score of 400 than it is with a group who averaged 600 at entrance.

- Fourth, some way must found to assess the accomplishments of graduating students and to express the results in terms that have meaning from year to year or across departments or, in some cases, across institutions. Faculty grades are too variable over time and departments to do the job. But if we don't want to rely on extrinsic indicators such as admission to graduate school, what can be done?

I propose one approach for consideration. To apply it, you need to preserve benchmark samples of actual student work, such as a set of senior theses (representing honors work, barely acceptable work, and three or four points in between), together with a record of what proportion of the senior class was able to reach each level in 1987. That will make a fascinating basis for comparison if you assign the same kind of senior exercises in 1990. It will be equally fascinating in 1987 if done cooperatively by a few interested colleges, using the SAT baseline data to monitor for any major differences between institutions in ability at entrance. The comparative performance of students who have come through different institutional experiences as preparation for a common senior-year exercise can be a rich source of data about the objectives that are or are not being achieved on one's own campus, provided some form of benchmark performances are used as markers in common across the colleges. This is a task for cooperative research among like-minded institutions.

- Fifth, the data must be wanted by somebody for some specific purpose. This fifth axiom might well have been the first. Traditionally, assessment data have been used mainly by people interested in institutional research. Fine. But the information is not going to have much immediate effect unless it is made vital to two other groups: first, the faculty and second, the administrators and policy makers. The latter group are especially important when the assessment data refer to whole classes or colleges and when the results can be related to policy decisions, which usually involve choices in the allocation of scarce resources among alternative uses.

Most sets of institutional data that have been collected in the past have been designed to answer broad retrospective questions about, for example, trends in student interests and values, or in the opinions of students, faculty, or alumni, rather than to serve as a basis for hard institutional decisions that are being contemplated for the future. When assessment is seen as a vital tool for academic management at both the classroom and the institutional level, it will be taken seriously. That won't happen until the assessment techniques are designed specifically to yield the kinds of information that somebody is depending on to illuminate issues and decisions for which that person bears direct responsibility.

Improving Educational Practice through Studying Changes in Student Progress

Traditionally we have evaluated institutions and programs by counting institutional resources, such as buildings, faculty ratios, or Ph.D.s. We went on to a faith in evaluation through the examination of processes, such as the breadth and depth of the curriculum. Now the pendulum—or at least the rhetoric—is swinging to using the assessment of student outcomes as the basis for evaluation. This is a healthy swing, in my judgment, which can be made even more significant if the focus comes to rest on student *progress* as well as terminal status. And finally, the new emphasis on student growth can play a dynamic role if we see that the road to further improvement is through discovering how changes that we choose to make in the traditional criteria of resources and processes result in changes in student progress, and if we use the information to improve practice. Then we will have closed the loop and taken an important step toward achieving the ideal of the self-improving institution.

7